Improved accuracy and precision of tracer kinetic parameters by joint fitting to variable flip angle and dynamic contrast enhanced MRI data

Improved accuracy and precision of tracer kinetic parameters by joint fitting to variable flip angle and dynamic contrast enhanced MRI data

Journal:	*Magnetic Resonance in Medicine*
Manuscript ID	MRM-15-16087.R3
Wiley - Manuscript type:	Full Paper
Date Submitted by the Author:	n/a
Complete List of Authors:	Dickie, Ben; The Christie NHS Foundation Trust, Christie Medical Physics and Engineering Banerji, Anita; University of Manchester, Institute of Population Health Kershaw, Lucy; The Christie NHS Foundation Trust, Christie Medical Physics and Engineering Choudhury, Ananya; The Christie NHS Foundation Trust, Department of Clinical Oncology McPartlin, Andrew; The Christie NHS Foundation Trust, Department of Clinical Oncology West, Catharine; University of Manchester, Institute of Cancer Sciences Rose, Chris; The University of Manchester, Institute of Population Health
Research Type:	Image processing/Image analysis < Theoretical < Technical Research, Mathematical modeling < Theoretical < Technical Research
Research Focus:	Cancer

Improved accuracy and precision of tracer kinetic parameters by joint fitting to variable flip angle and dynamic contrast enhanced MRI data

Ben R. Dickie[1,3], Anita Banerji[2], Lucy E. Kershaw[1,3], Andrew McPartlin[3,4], Ananya Choudhury,[3,4] Catharine M. West[3], Chris J. Rose[2]

September 18, 2015

* Corresponding author: ben.dickie@postgrad.manchester.ac.uk

1. Christie Medical Physics and Engineering, The Christie NHS Foundation Trust, Manchester, United Kingdom
2. Institute of Population Health, Centre for Imaging Sciences, Manchester Academic Health Science Centre, The University of Manchester, Manchester, United Kingdom
3. Institute of Cancer Sciences, Manchester Academic Health Science Centre, The University of Manchester, Manchester, United Kingdom
4. Department of Clinical Oncology, The Christie NHS Foundation Trust, Manchester, United Kingdom

Running head: Improved accuracy and precision of DCE-MRI tracer kinetic parameters

Approx Word count: 4500

Abstract

Purpose:

To improve the accuracy and precision of tracer kinetic model parameter estimates for use in dynamic contrast enhanced (DCE) MRI studies of solid tumors.

Theory:

Quantitative DCE-MRI requires an estimate of pre-contrast T_1, which is obtained prior to fitting a tracer kinetic model. As T_1 mapping and tracer kinetic signal models are both a function of pre-contrast T_1 it was hypothesized that its joint estimation would improve the accuracy and precision of both pre-contrast T_1 and tracer kinetic model parameters.

Methods:

Accuracy and/or precision of two-compartment exchange model (2CXM) parameters were evaluated for standard and joint fitting methods in well-controlled synthetic data and for 36 bladder cancer patients. Methods were compared under a number of experimental conditions.

Results:

In synthetic data, joint estimation led to statistically significant improvements in the accuracy of estimated parameters in 30 of 42 conditions (improvements between 1.8% and 49%). Reduced accuracy was observed in 7 of the remaining 12 conditions. Significant improvements in precision were observed in 35 of 42 conditions (between 4.7% and 50%). In clinical data, significant improvements in precision were observed in 18 of 21 conditions (between 4.6% and 38%).

Conclusion:

Accuracy and precision of DCE-MRI parameter estimates are improved when signal models are fit jointly rather than sequentially.

Key words: DCE-MRI, tracer kinetic modeling, joint estimation, precision, accuracy, heterogeneity, two-compartment exchange model

Introduction

The goal of quantitative dynamic contrast enhanced (DCE) MRI is to estimate tracer kinetic model parameters for a tissue of interest. To fit a tracer kinetic model to DCE-MRI data, MR signal intensity must first be converted to contrast agent concentration, a process which requires an estimate of pre-contrast T_1 ($T_{1,0}$) (1). Errors in $T_{1,0}$ will propagate through to errors in estimates of contrast agent concentration, eventually affecting the tracer kinetic parameters of interest (2, 3). To be clinically useful, acquisition times for T_1 mapping data should be of the order of seconds-minutes, discounting the use of gold standard methods (e.g. multi-point inversion recovery spin-echo sequences). Faster techniques using gradient echo turboFLASH sequences have been proposed (4, 5) and T_1 estimates using these methods agree well with their spin-echo counterparts (4), however acquisition times are still too long in some applications, especially if other quantitative imaging acquisitions such as diffusion weighted MRI are required. Many DCE-MRI studies within the last 15 years (6, 7, 8, 9) have opted to perform T_1 mapping using the spoiled gradient recalled echo (SPGR) variable flip angle (VFA) technique (10), which is less accurate and precise than multi-point methods (11, 12), but can provide the required coverage in a matter of seconds.

Image noise contributes significant error to VFA T_1 estimates (13). Noise can be minimized by acquiring multiple signal averages, however this increases acquisition time reducing the benefits of the VFA method compared to more accurate T_1 mapping techniques. In the context of DCE-MRI, multiple dynamic contrast enhanced images are acquired following the VFA data. Often the same sequences are used for both acquisitions, however information about $T_{1,0}$ within the dynamic acquisitions is then ignored. Since both VFA and dynamic signal models are a function of $T_{1,0}$, such information could theoretically be included during its estimation by jointly fitting signal models to VFA and dynamic data. Work outside the field of MRI showed that joint fitting of signal models which share parameters can improve the accuracy and precision of parameter estimates compared to when models are fit separately (14, 15).

Recognizing that $T_{1,0}$ is a common parameter to both VFA and dynamic signal models, we hypothesized that joint fitting of these models would improve the accuracy and precision of $T_{1,0}$ and tracer kinetic parameter estimates. This paper describes the theory behind the standard sequential and proposed joint estimation approaches. The hypothesis is then tested using the two-compartment exchange model (2CXM) in well-controlled synthetic and clinical data. Additional experiments are presented in supporting materials.

Theory

Sequential estimation

This section describes the standard approach to estimate tracer kinetic parameters from SPGR VFA and DCE-MRI data, and suggestions are made as to why it may be suboptimal. Tracer kinetic parameters are estimated by fitting a model (e.g. 2CXM (16)) to concentration time courses. Since contrast agent concentration cannot be measured directly using MRI, it must be inferred from measured signal time courses using an estimate of $T_{1,0}$. The tracer kinetic model can also be used to compute contrast agent concentrations, which can be converted to idealized signal values (using the estimated $T_{1,0}$) and fitted to measured dynamic signal. This second approach is described in detail below.

In general, measured MR signal magnitude, y, at an arbitrary voxel can be modeled as:

$$y = s + \epsilon \tag{1}$$

where s is the underlying noise-free signal and ϵ is an independent and identically distributed (i.i.d) random variable, modeling image noise. For image data acquired using an SPGR sequence, the underlying noise-free signal can be modeled as (17):

$$s = \frac{M_0 \sin\theta(1 - e^{-\frac{TR}{T_1}})}{1 - \cos(\theta)e^{-\frac{TR}{T_1}}} \tag{2}$$

where TR is the repetition time, θ is the flip angle, and T_1 and M_0 are the spin-lattice relaxation time and equilibrium longitudinal magnetization at the voxel respectively. Typically the echo time, TE, is kept short such that signal decay due to T_2^* effects can be ignored. To differentiate between VFA and dynamic flip angles, we use vector $\boldsymbol{\theta_v} = [\theta_{v_1}, \theta_{v_2}, \theta_{v_3} \ldots \theta_{v_N}]$ and scalar θ_d respectively, where N is the number of distinct flip angles in the VFA set. The equilibrium longitudinal magnetization (M_0) depends on proton density and receive gain. In general, M_0 may differ between corresponding voxels in VFA and dynamic images and therefore $M_{0,v}$ and $M_{0,d}$ are defined respectively.

Assuming a signal to noise ratio > 3, noise is well approximated as Gaussian ($\epsilon \sim \mathrm{N}(0, \eta)$), where η is the standard deviation of the noise present in the images (18). Since noise may differ between VFA and dynamic images we define η_v and η_d respectively. Substituting $T_1 = T_{1,0}$ and $M_0 = M_{0,v}$ into Eqn 2, estimates of $\hat{T}_{1,0}$ and $\hat{M}_{0,v}$ are obtained from the VFA signal by maximizing the following log-likelihood function with respect to $T_{1,0}$ and $M_{0,v}$:

$$\log \mathcal{L}_v = -\frac{N \log(2\pi\eta_v^2)}{2} - \sum_{i=1}^{N} \frac{(y(\theta_{v_i}) - s(\theta_{v_i}))^2}{2\eta_v^2} \tag{3}$$

where $y(\theta_{v_i})$ is the measured VFA signal at the voxel for flip angle θ_{v_i} and N is the total number of distinct flip angles within the VFA image set. In the case of Gaussian errors, maximization of log-likelihood functions is equivalent to minimization of the sum of squared residuals.

Next, $\hat{T}_{1,0}$ is used in conjunction with an estimate of the mean pre-contrast dynamic signal, $\hat{s}_{\text{pre}}$, to obtain an estimate the equilibrium longitudinal magnetization at the voxel within the dynamic images, $\hat{M}_{0,d}$. Eqn 2 can be restated as:

$$\hat{M}_{0,d} = \frac{\hat{s}_{\text{pre}}(1 - \cos(\theta_d)e^{-\frac{TR}{\hat{T}_{1,0}}})}{\sin\theta_d(1 - e^{-\frac{TR}{\hat{T}_{1,0}}})} \tag{4}$$

where,

$$\hat{s}_{\text{pre}} = \frac{1}{n_{\text{pre}}}\sum_{j=1}^{n_{\text{pre}}} y(t_j) \tag{5}$$

where $y(t_j)$ is the measured dynamic signal at acquisition time t_j and n_{pre} is the number of pre-contrast dynamic time points.

After arrival of contrast agent at the voxel, a reduction in the T_1 relaxation time is observed. Assuming the fast exchange limit for water exchange, the T_1 relaxation rate at dynamic acquisition time t_j is given by:

$$\frac{1}{T_1(t_j)} = rC(t_j) + \frac{1}{\hat{T}_{1,0}} \tag{6}$$

where r is the T_1 relaxivity of the contrast agent and $C(t_j)$ is the contrast agent concentration at time t_j. Substituting Eqn 6 into Eqn 2, with $M_0 = \hat{M}_{0,d}$ gives:

$$s(t_j) = \frac{\hat{M}_{0,d}\sin\theta_d(1 - e^{-TR(rC(t_j)+\frac{1}{\hat{T}_{1,0}})})}{1 - \cos(\theta_d)e^{-TR(rC(t_j)+\frac{1}{\hat{T}_{1,0}})}} \tag{7}$$

The concentration of contrast agent at the voxel, $C(t_j)$, can be modeled as a convolution between the tissue's impulse response function (IRF) and the tissue's arterial input function (AIF). For the 2CXM, the IRF is described by 4 microvascular parameters, $\boldsymbol{p} = [F_{\text{p}}, F_{\text{E}}, v_{\text{p}}, v_{\text{e}}]$, where F_{p} is the plasma flow, F_{E} is the exchange flow, v_{p} is the plasma volume and v_{e} is the interstitial volume (19). Units for 2CXM parameters are shown in Table 1. The tracer kinetic parameters $\boldsymbol{p}$ are estimated from the dynamic signal by maximizing the following log-likelihood function with respect to $\boldsymbol{p}$:

$$\log\mathcal{L}_d = -\frac{n\log(2\pi\eta_d^2)}{2} - \sum_{j=1}^{n}\frac{(y(t_j) - s(t_j))^2}{2\eta_d^2} \tag{8}$$

where n is the number of dynamic time points.

Tracer kinetic analysis using this sequential approach is well-established but has three main shortcomings:

1. While dynamic fits resulting from sequential estimation may appear to fit the data well, error in $\hat{T}_{1,0}$ and $\hat{M}_{0,d}$ will cause the underlying dynamic likelihood function (Eqn. 8) to be erroneous, causing the tracer kinetic parameters to be erroneous.
2. Substitution of VFA $\hat{T}_{1,0}$ into the dynamic signal model as a fixed parameter is statistically inefficient because potentially useful $T_{1,0}$ information within the dynamic images is ignored.
3. If $M_{0,v} = M_{0,d}$, statistical power is lost by making two estimates ($M_{0,v}$ and $M_{0,d}$) of the same underlying parameter.

Joint estimation

The problems identified above can be addressed by estimating $T_{1,0}$, $M_{0,v}$, $M_{0,d}$, and $\boldsymbol{p}$ jointly rather than sequentially. In this framework, $T_{1,0}$ information within both the VFA and dynamic images is allowed to contribute to the estimate of $T_{1,0}$. Also, when $M_{0,v} = M_{0,d}$, a single equilibrium longitudinal magnetization parameter can be estimated at each voxel, instead of two. To facilitate exposition, this single estimate is called $M_{0,d}$, even though it is estimated jointly from both VFA and dynamic images. Joint estimation can be performed by simply maximizing a log-likelihood function resulting from the sum of the VFA and dynamic log-likelihood functions used for sequential estimation:

$$[\hat{T}_{1,0},\ \hat{M}_{0,d},\ \hat{\boldsymbol{p}}] = \arg \max_{T_{1,0},\ M_{0,d},\ \boldsymbol{p}} (\log \mathcal{L}_v + \log \mathcal{L}_d) \quad (9)$$

Methods

Well-controlled synthetic data were used to test the null hypothesis of no difference in 2CXM parameter accuracy between sequential and joint estimation. Well-controlled synthetic data and clinical data from 36 bladder cancer patients (7) were used to test the null hypothesis of no difference in 2CXM parameter precision between sequential and joint estimation. Accuracy of tracer kinetic parameters could not be assessed in the clinical data because of the lack of ground truth. Accuracy of sequential and joint $T_{1,0}$ estimates were evaluated in 1532 voxels from a clinical prostate cancer study by comparing estimates to independent inversion-recovery turbo-field echo (IR-TFE) measurements.

While it is hypothesized that joint estimation will improve the accuracy and precision of all estimated parameters, joint estimation using highly erroneous dynamic data may be expected to result in poorer estimates. Therefore, the effect of three sources of systematic error on sequential and joint parameter estimates were investigated: errors due to B_1 field inhomogeneity, errors due to underestimation of the

AIF and errors due to overestimation of the AIF. The effects of B_1 field inhomogeneity on sequential and joint estimates were studied only in the synthetic data as B_1 homogeneity could not be manipulated retrospectively in the clinical dataset. While it may be possible in a future study to map the B_1 field using joint estimation, it was not investigated within the current paper.

A highly-realistic publicly-available software phantom generator (20) was used within a Monte Carlo framework to simulate and analyze 100 liver tumor VFA images and DCE image series for each experimental condition. For the main clinical experiment, VFA and DCE images from 36 bladder tumors were analyzed within a residual bootstrapping framework. A random sample of voxels were selected from the synthetic and clinical tumors and data from those voxels were used in each experiment. In each experimental condition, Monte Carlo or bootstrap iteration, and each voxel in the sample, we estimated $T_{1,0}$, $M_{0,v}$, $M_{0,d}$, t_0, F_p, F_E, v_p, and v_e using sequential and joint estimation (t_0 is the offset time between bolus arrival at the arterial sampling point and at the tissue). To evaluate accuracy, the deviation between an estimate and its correct value was computed. For precision, the deviation between an estimate and its expected value was computed. Multivariate linear regression was used to estimate the difference in accuracy and precision between joint and sequential estimates attributable to each experimental condition. Point estimates and Bonferroni corrected 95% confidence intervals on the percentage improvement in accuracy and precision due to joint estimation were tabulated for each parameter and experimental condition. To determine the number of voxels to include in the experiments, an *a priori* sample size calculation was made using G*Power (version 3.1.9.2.). Parameter maps were constructed for the synthetic tumor and two representative tumors from the clinical bladder data. For the prostate experiment, the null hypothesis of no difference in accuracy between sequential and joint estimates of $T_{1,0}$ compared to IR-TFE measurements was tested (see supporting materials). Software to run the experiments is available at http://github.com/MRdep/Joint-fitting (21).

Synthetic data

Phantom anatomy was based on organ masks defined on end-exhale DCE-CT data from a single individual with a liver tumor (tumor volume = 28447 mm^3, 2040 voxels). Microvascular heterogeneity was simulated by segmenting the tumor into two distinct regions representing a highly perfused, highly vascularized rim and a poorly perfused, poorly vascularized core (Fig 1). In each tumor voxel, contrast agent kinetics were simulated using the 2CXM. The AIF used to generate signal-time curves was measured from a randomly selected patient in the clinical bladder cohort (Fig 2a, accurate AIF). Sampling distributions for ground truth $T_{1,0}$, $M_{0,v}$, $M_{0,d}$, t_0, and 2CXM parameters are shown in Table 1. VFA and dynamic image acquisition parameters were chosen to match the clinical bladder protocol: field of view (FOV) of 240 × 320 × 80 mm^3; voxel size of 1.67 × 1.67 × 5.00 mm^3; TR/TE = 3.2/1.2 ms; dynamic temporal resolution of 2.5 s; and dynamic scan duration of 4 minutes. Source code to generate synthetic images are available at http://www.qbi-lab.org/software.php.

Insert figure 1 about here.

In the homogeneous B_1 field condition, images were created using spatially uniform flip angles. These were set equal to those prescribed in the clinical protocol: VFA flip angles of 5°, 10° and 35° and a dynamic flip angle of 25°. In the inhomogeneous B_1 field condition, images were created by randomly varying the flip angle error across the imaging volume between 50% and 150% of the prescribed flip angles. This range was chosen to represent a worst case scenario; B_1 field inhomogeneities at 1.5 T and 3 T would typically not be this severe (22, 23, 24).

Insert table 1 about here.

Clinical data

Retrospective analysis was performed on VFA and DCE-MRI scans from 36 patients with muscle invasive bladder cancer (age range 45–74 years, mean 63 years). All patients gave written informed consent and approval was obtained from the local research ethics committee. Scanning was performed on a 1.5 T Siemens Magnetom Avanto MR scanner (Siemens Medical Solutions, Erlangen, Germany). A 2D T_2-weighted turbo spin echo scan (TR/TE = 4000/99 ms, NSA = 1) covering the same FOV as the subsequent VFA and dynamic scans but with improved spatial resolution (voxel size of 0.63 × 0.63 × 5 mm^3) was used for the purpose of defining a tumor ROI. For VFA and dynamic acquisitions, a 3D T_1-weighted volumetric interpolated breath-hold examination (VIBE) sequence was employed with the same scan parameters as the synthetic data, except with a SENSE factor of 2. VFA imaging was performed with 5 signal averages. No averaging was performed during dynamic imaging. All images were acquired in the transverse plane with the FOV encompassing the whole bladder (7).

Gadolinium-based contrast agent (Magnevist, Bayer-Schering Pharma AG, Berlin, Germany) was injected as a 0.1 mmol/kg bolus with a power injector through a cannula placed in the antecubital vein. The injection was administered 15 s into the dynamic acquisition at 3 ml/s, and was followed by a 20 ml saline flush.

Tumor ROIs were delineated by two radiologists in consensus (S.B. and B.C. with 15 and 23 years experience respectively, see acknowledgments) and transferred via down-sampling to the VFA and dynamic images. AIFs were extracted using a semi-automatic procedure described previously (7). Signal from the arterial ROI was converted to plasma contrast agent concentration using the SPGR equation assuming a literature value for blood $T_{1,0}$ of 1480 ms (25) and hematocrit of 0.42.

AIF errors

Direct measurement of AIFs from DCE-MRI data is difficult and errors can arise due to the presence of inflow effects, partial volume effects, T_2^* decay and water exchange effects. To assess the impact of

AIF error on sequential and joint parameter estimates, the true AIF (in the synthetic experiment) or measured AIF (in the main clinical experiment) were scaled by factors of 0.5 and 1.5, leading to under and over-estimated AIFs respectively. These scaling factors were based on errors in peak concentration previously reported in phantom, pre-clinical and clinical data at 1.5 T due to partial volume effects, inflow effects and T_2^* decay (26, 22, 27, 28, 29).

Monte Carlo and residual bootstrap analyses

Monte Carlo analysis (30) was performed on the synthetic data to facilitate the use of idealized distributions for random measurement error processes. Random measurement error was modeled as samples from a zero mean Gaussian distribution. The variance of the distributions for the VFA (η_v^2) and dynamic data (η_d^2) were chosen to give SNRs of $5\sqrt{5}$ and 5 respectively to mimic that expected in the clinical bladder data. Residual bootstrap analysis (31) was performed in the main clinical experiment to facilitate the use of natural distributions for random measurement error processes, without requiring us to assume errors follow idealized distributions.

Model fitting

All model fitting was performed in IDL 8.2.2 (Exelis Visual Information Solutions, Boulder, Colorado, USA) using the function 'mpcurvefit'. Initial estimates of parameters were set at: $T_{1,0} = 500$ ms, $M_{0,v} = 5000$ a.u., $M_{0,d} = 5000$ a.u., $F_\mathrm{p} = 0.5$ ml min^{-1} ml^{-1}, $F_\mathrm{E} = 0.5$ ml min^{-1} ml^{-1} and $v_\mathrm{e} = 0.2$ ml ml^{-1}. An initial estimate for the offset time, $t_{0,i}$ (min) was calculated by fitting the Tofts model (1) to the initial third of the dynamic time series, with t_0 as a free parameter. For sequential estimation, the dynamic signal model was fitted with F_p, F_E, v_e and t_0 as free parameters. $T_{1,0}$ and $M_{0,d}$ were fixed to estimates obtained from VFA and pre-contrast dynamic data. For joint estimation, VFA and dynamic signal models were fitted jointly with $T_{1,0}$, $M_{0,v}$, $M_{0,d}$, F_p, F_E, v_e and t_0 as free parameters, with $M_{0,v}$ constrained to be equal to $M_{0,d}$ (supporting materials describe an experiment showing this is a reasonable assumption for our clinical protocols). For both sequential and joint estimation, v_p was fixed and incremented from 0 - 1 in 0.01 steps over the course of 100 repeated fits. The fit giving the maximum log-likelihood was chosen. In all optimizations, the following parameter constraints were imposed: $0 < T_{1,0} < 5$ s, $0 < M_{0,v} < 40000$ a.u., $0 < M_{0,d} < 40000$ a.u. $F_\mathrm{p} > 0$ ml min^{-1} ml^{-1}, $F_\mathrm{E} > 0$ ml min^{-1} ml^{-1}, $0 < v_\mathrm{e} < 1$ ml ml^{-1}, $t_{0,i}$ - $\frac{5}{60} < t_0 < t_{0,i} + \frac{5}{60}$ min. Convolutions were computed using trapezoidal integration. In both the synthetic and clinical experiments, η_v was set equal to $\eta_d/\sqrt{5}$ to account for differences in the expected noise.

Accuracy and precision

Accuracy was defined as the absolute relative difference between an estimate and its correct value, $\lambda = |(\hat{x} - x)/x|$. To characterize the improvement in accuracy of joint estimation (λ^J) over sequential estimation (λ^S) we defined the ratio: $\Lambda = \frac{\lambda^J}{\lambda^S}$, where the percentage improvement in accuracy is $(1 - \Lambda) \times 100\%$.

Precision was defined as the absolute relative difference between an estimate and its *expected value*, $\omega = |(\hat{x} - \bar{x})/\bar{x}|$. Within each experimental condition, the expected value $\bar{x}$ at each voxel was estimated by taking the mean over Monte Carlo or residual bootstrapping iterations. To characterize the improvement in precision of joint estimation (ω^J) over sequential estimation (ω^S) we defined the ratio: $\Omega = \frac{\omega^J}{\omega^S}$, where the percentage improvement in precision is $(1 - \Omega) \times 100\%$.

Statistical analysis

Three multivariate linear regressions were performed to test hypotheses about accuracy for the synthetic data, precision for the synthetic data and precision for the clinical data. Each multivariate linear model took the following form:

$$z_k = \beta_0 + \beta_1 X_k^U + \beta_2 X_k^O + \beta_3 X_k^{B_1} + \epsilon_k \tag{10}$$

The subscript k indexes Monte Carlo or bootstrap iteration over all experimental conditions. On the left hand side, z_k is $\log \Lambda$ or $\log \Omega$ for accuracy and precision respectively. The dependent variable was defined on the log scale to improve the normality of the residuals; this corresponds to the difference in $\log \lambda$ and $\log \omega$ (i.e. $\log \lambda^J - \log \lambda^S$ and $\log \omega^J - \log \omega^S$). On the right hand side, X_k^U takes a value of 1 in the under-estimated AIF condition and 0 otherwise; X_k^O takes a value of 1 in the over-estimated AIF condition and 0 otherwise; $X_k^{B_1}$ takes a value of 1 in the inhomogeneous B_1 field condition and 0 otherwise; and ϵ_k is residual error. The model coefficient β_0 corresponds to the sample mean of z under the reference conditions (i.e. accurate/measured AIF and homogeneous B_1 field); β_1, β_2 and β_3 quantify the residual mean z attributable to the under-estimated AIF condition, over-estimated AIF condition and inhomogeneous B_1 field condition respectively. In the clinical data, $\beta_3 = 0$ because the B_1 field could not be manipulated. Point estimates and 95% confidence intervals on the percentage improvement in accuracy and precision were computed by transforming from the log space to the original data space. Statistical analysis was performed in R (Version 3.1, R Foundation for Statistical Computing, Vienna, Austria).

Sample size

Sample size calculations were based on detecting a medium effect size (Cohen's $f^2 = 0.15$) at statistical power of 95% and significance level of 0.05, Bonferroni-corrected for the total number of inferences made. For each of the 3 experimental conditions in the main clinical experiment, we randomly sampled 14 voxels from each of the 36 tumors (504 voxels in total), and performed 100 iterations of the residual bootstrap analysis on each, for a total sample size of just over 1.5 million. For each of the 6 conditions of the synthetic experiment, we randomly sampled 504 voxels from the synthetic tumor and performed 100 iterations of the Monte Carlo analysis on each, for a total sample size of just over 3 million. The same voxels were analyzed under each experimental condition, resulting in a repeated measures design.

Results

Figure 1 shows example synthetic images created using the software phantom generator. Figures 2a and 2b show AIFs used in the synthetic experiment and for one example patient in the main clinical experiment. The accurate AIF in Figure 2a was used to generate the DCE-MRI time courses in the synthetic tumor. Figures 2c and 2d show Monte Carlo and residual bootstrap fits ($n = 100$) for example voxels in the synthetic and clinical experiments. Figures 2e and 2f show corresponding density estimates obtained from Monte Carlo and residual bootstrap experiments shown in 2c and 2d. A number of key differences between sequential and joint fits were observed. For the synthetic data, joint VFA fits showed less variability in shape compared to sequential fits. For all fits to dynamic data, modeled pre-contrast signal was less variable with joint estimation. Poor fits obtained with sequential estimation in the under-estimated AIF condition were not observed with joint estimation. In contrast to the synthetic experiment, joint fits to clinical VFA data showed greater variability than sequential fits. Regardless of this, the effect of joint estimation in the dynamic data mirrored that observed for the synthetic experiment.

In the synthetic data, parameter densities were narrower for sequential estimates of $T_{1,0}$, suggesting joint T_1 estimates had lower absolute precision. However, since joint $T_{1,0}$ estimates were also shifted to higher values (which were also more accurate), relative precision was increased compared with sequential estimates. In the clinical data, the shift to higher values was not so profound, and relative precision in $T_{1,0}$ was therefore degraded when using joint estimation. For 2CXM parameters, joint estimation led to narrower distributions which were also shifted closer to ground truth, reflecting increased precision and accuracy. Similar benefits in 2CXM parameter precision were observed in the clinical data.

Insert figure 2 about here.

Tables 2, 3, and 4 show results from the multivariate linear regression models. Tables 2 and 3 show average percentage improvements in accuracy and precision due to joint estimation in the synthetic data. Table 4 shows average percentage improvement in precision due to joint estimation in the main clinical

experiment. In synthetic data, statistically significant improvements in accuracy were seen in 30 out of 42 cases (a case represents a unique parameter/experimental condition combination). Improvements for $T_{1,0}$, $M_{0,d}$ and t_0 were between 7.7% and 49%. Improvements for the tracer kinetic parameters were between 1.8% and 21%. Statistically significant detriments were observed in 7 out of 42 cases, mainly in the over-estimated AIF condition (between -5.4% and -20%), and in the case of v_p under the inhomogeneous B_1 field condition (between -3.8% and -20%). Statistically significant improvements in precision were seen in 35 of 42 cases (between 4.7% and 50%). No statistically significant decreases in precision were observed. In clinical data, statistically significant improvements in precision were observed in 18 of 21 cases (between 4.6% and 38%). Statistically significant decreases in precision were observed for $T_{1,0}$ in 2 of the remaining 3 cases (between -4.9% and -8.5%).

Insert table 2 about here.

Insert table 3 about here.

Insert table 4 about here.

Figure 3 shows $T_{1,0}$, $M_{0,d}$, F_p, v_p, F_E and v_e maps for the synthetic experiment. Differences in joint and sequential $T_{1,0}$ maps were difficult to determine visually, however $M_{0,d}$ maps obtained using joint estimation showed less speckle and better agreement with ground truth compared with sequential maps. Under the homogeneous B_1 field condition, the rim-core boundary present in the ground truth F_p map could be clearly identified for both sequential and joint fitting methods. Under the inhomogeneous B_1 field condition, the rim-core boundary could not be easily distinguished in the sequential maps, but could be clearly identified in the joint estimation maps. Joint estimation also appeared to reduce the variability in estimates of v_p (less speckle).

Insert figure 3 about here.

Figure 4 shows $T_{1,0}$, $M_{0,d}$, F_p, v_p, F_E and v_e maps for two example tumors from the main clinical experiment. In tumor 1, joint F_p maps appeared less sensitive to error in the AIF, and joint v_p maps were less speckled than their sequential estimation counterparts. In example tumor 2, the number of spurious v_p, F_E, and v_e values observed near the tumor center were reduced with joint estimation.

Insert figure 4 about here.

Supporting Figure 1 shows sequential and joint estimates of $T_{1,0}$ compared against independent IR-TFE $T_{1,0}$ measurements in 1532 voxels from a clinical prostate cancer study. On average, joint estimates of $T_{1,0}$ laid closer to IR-TFE measurements compared to sequential estimates. Using IR-TFE measurements as an independent gold-standard, estimates of $T_{1,0}$ obtained using joint estimation were significantly more accurate than sequential estimates (improvement in mean relative error of 20%, $P < 0.0001$).

Discussion

A novel method for improving the accuracy and precision of tracer kinetic parameters was proposed and evaluated. The method recognizes that signal models used to describe VFA and DCE-MRI data share parameters, and utilizes this shared information by jointly fitting models to the observed data.

Results from the synthetic experiment show that joint estimation leads to large improvements in accuracy and precision of tracer kinetic parameter estimates under a range of experimental conditions. Improvements were likely caused by increased accuracy and precision of $T_{1,0}$ and $M_{0,d}$, a consequence of including additional pre-contrast T_1 and equilibrium longitudinal magnetization information during fitting. With more data contributing to estimates of $T_{1,0}$ and $M_{0,d}$, random signal errors were thought to have a smaller effect on the log-likelihood function, therefore reducing the effect of noise on parameter estimates.

During joint estimation it was expected that errors in the dynamic data may contribute additional error to $T_{1,0}$, which would not occur during sequential estimation. This was tested by simulating a number of error conditions in synthetic data and assessing accuracy of both methods under each condition. A comparison was also made between sequential and joint estimates of $T_{1,0}$ and independent measurements obtained using an IR-TFE sequence in 1532 voxels from a clinical prostate cancer study (see supporting materials). In the synthetic data, accuracy of $T_{1,0}$ was improved with joint estimation under all error conditions, suggesting that inclusion of $T_{1,0}$ information stored within the dynamic images outweighs the potentially detrimental effect of dynamic signal errors. In the prostate data, joint estimation led to improved accuracy in $T_{1,0}$ comparable to that observed in the synthetic data (taking IR-TFE measurements as gold standard, improvements were 20% compared to 10-22% in the synthetic data).

While $T_{1,0}$ accuracy was improved with joint estimation in a range of simulated imaging scenarios, these did not always translate to improvements in accuracy of tracer kinetic parameters, especially when the AIF was overestimated. However, in the case of accurate and underestimated AIFs, joint fitting led to significant improvements in the accuracy of nearly all parameters. While degradations in the accuracy of tracer kinetic parameters are not ideal, inaccuracy can in many cases be compensated via calibration. Precision, on the other hand, is a stochastic characteristic and can only be increased by reducing the variability of the measurement process. We therefore stress the importance of the significant improvements in precision of tracer kinetic parameters observed with joint estimation. Increased precision can lead to greater statistical power when aiming to detect longitudinal changes in a parameter within the same patient over time (e.g. in a clinical trial setting) or when detecting differences in DCE-MRI parameters between patients (e.g. for prediction of response to therapy). It also improves our ability to distinguish tumor subregions, useful for analyzing tumor heterogeneity (32), as demonstrated in Figure 3.

Observed improvements in precision were similar in the synthetic and clinical bladder experiment however there were some key differences. For example, v_e precision was improved in the bladder data but not the synthetic data. This was probably because absolute precision of v_e was higher in the synthetic data,

making improvements with joint estimation difficult to achieve. Poorer absolute precision in v_e within the bladder data may have been caused by tumor motion or reduced tracer back flux from the interstitial space during imaging (compared to that simulated in the synthetic data) (33). Differences in the improvement of $T_{1,0}$ precision between synthetic and bladder experiments were also observed. In the synthetic experiment, $T_{1,0}$ precision was improved by around 20%, however in the clinical data $T_{1,0}$ precision was degraded. The latter observation is likely due to underestimation of the noise present in the clinical dynamic images (possibly because tumor motion was not considered during noise estimation), which led to overweighting of these data points within the joint log-likelihood function. Regardless, improvements were still observed for $M_{0,d}$, leading to significant improvements in the precision of 2CXM parameters.

This study has the following limitations. We could not assess accuracy of tracer kinetic parameters in clinical data due to lack of ground truth, and rely on simulations only. While sequential and joint estimation were evaluated across a wide range of simulated experimental conditions (B_1 field errors and AIF errors), the effect of patient motion and image artifacts were not considered. Furthermore, while linear scaling of the AIF may have accurately simulated first order perturbations associated with inflow, partial volume effects, T_2^* signal decay and water exchange effects, changes to arterial concentration time curves due to these sources of error are likely to be non-linear. Further work should aim to more accurately simulate specific AIF errors as well as patient motion, and assess the effect of such errors on jointly estimated parameters. Future work could also study the possibility of mapping the B_1 field with joint estimation by including flip angle error as a free parameter during fitting. The current study evaluated the benefits of joint estimation for a single model only, but in heterogeneous lesions the optimal tracer kinetic model may vary from voxel to voxel. Since joint fits to VFA data depend on fits to the dynamic data, joint estimation is likely to provide benefits only if the tracer kinetic model is valid for the tissue of interest. Last, although Monte Carlo (30) and residual bootstrapping methods (31) are well-accepted techniques for evaluating accuracy and precision, they do not enable differences in parameter estimates caused by variations in patient positioning, scanner calibration, coil positioning, AIF selection etc. to be taken into account. To evaluate these effects, the reproducibility of jointly estimated parameters should be compared to sequential estimates in a clinical trial setting.

The hypothesis underlying the work of this paper was that joint estimation would improve the accuracy and precision of tracer kinetic parameters by considering variables common to T_1 mapping and dynamic signal models. This hypothesis was supported by showing moderate to large statistically significant improvements in accuracy (1.8% to 49%) and precision (4.7% to 50%) for most model parameters in most experimental conditions. It is therefore recommended that investigators consider using joint estimation instead of sequential estimation, particularly given that joint estimation is straightforward to implement and requires no or little modification of commonly-used DCE-MRI protocols.

Acknowledgments

B.R.D. is funded by Christie Medical Physics and Engineering. Thanks to Suzi Bonington (S.B.) and Bernadette Carrington (B.C.) of the Christie Hospital for outlining tumor regions of interest in the bladder patients. Thanks to Kristy Brock (University of Michigan, USA) for kindly sharing organ masks used in the phantom generator software.

Bibliography

[1] Tofts PS. Modeling tracer kinetics in dynamic Gd-DTPA MR imaging. Journal of Magnetic Resonance Imaging 1997;7:91–101.

[2] Di Giovanni P, Azlan C, Ahearn T, Semple S, Gilbert F, Redpath T. The accuracy of pharmacokinetic parameter measurement in DCE-MRI of the breast at 3 T. Physics in Medicine and Biology 2010; 55:121.

[3] Garpebring A, Brynolfsson P, Yu J, Wirestam R, Johansson A, Asklund T, Karlsson M. Uncertainty estimation in dynamic contrast-enhanced MRI. Magnetic Resonance in Medicine 2013;69(4):992–1002.

[4] Blüml, Stefan and Schad, Lothar R and Stepanow, Boris and Lorenz, Walter J. Spin-lattice relaxation time measurement by means of a TurboFLASH technique. Magnetic Resonance in Medicine 1993; 30(3):289–295.

[5] Parker GJ, Baustert I, Tanner SF, Leach MO. Improving image quality and T1 measurements using saturation recovery turboFLASH with an approximate K-space normalisation filter. Magnetic Resonance Imaging 2000;18(2):157–167.

[6] O'Connor J, Rose C, Jackson A, *et al.*. DCE-MRI biomarkers of tumour heterogeneity predict CRC liver metastasis shrinkage following bevacizumab and FOLFOX-6. British Journal of Cancer 2011; 105(1):139–145.

[7] Donaldson SB, Bonington SC, Kershaw LE, Cowan R, Lyons J, Elliott T, Carrington BM. Dynamic contrast-enhanced MRI in patients with muscle-invasive transitional cell carcinoma of the bladder can distinguish between residual tumour and post-chemotherapy effect. European Journal of Radiology 2013;82:2161–2168.

[8] Buckley DL, Roberts C, Parker GJ, Logue JP, Hutchinson CE. Prostate Cancer: Evaluation of Vascular Characteristics with Dynamic Contrast-enhanced T1-weighted MR Imaging - Initial Experience. Radiology 2004;233(3):709–715.

[9] Ton NC, Parker GJ, Jackson A, *et al.*. Phase I evaluation of CDP791, a PEGylated di-Fab conjugate that binds vascular endothelial growth factor receptor 2. Clinical Cancer Research 2007;13(23):7113–7118.

[10] Fram EK, Herfkens RJ, Johnson GA, Glover GH, Karis JP, Shimakawa A, Perkins TG, Pelc NJ. Rapid calculation of T1 using variable flip angle gradient refocused imaging. Magnetic Resonance Imaging 1987;5:201–208.

[11] Andreisek, G and White, L and Yang, Y and Robinson, E and Cheng, HLM and Sussman, M. Delayed Gadolinium-enhanced MR Imaging of Articular Cartilage: Three-dimensional T1 Mapping with Variable Flip Angles and B1 Correction. Radiology 2009;252(3):865–873.

[12] Siversson C, Tiderius CJ, Neuman P, Dahlberg L, Svensson J. Repeatability of T1-quantification in dGEMRIC for three different acquisition techniques: Two-dimensional inversion recovery, three-dimensional look locker, and three-dimensional variable flip angle. Journal of Magnetic Resonance Imaging 2010;31(5):1203–1209.

[13] Cheng HLM, Wright GA. Rapid high-resolution T1 mapping by variable flip angles: Accurate and precise measurements in the presence of radiofrequency field inhomogeneity. Magnetic Resonance in Medicine 2006;55:566–574.

[14] Motulsky H, Christopoulos A. Fitting models to biological data using linear and nonlinear regression: a practical guide to curve fitting. Oxford University Press, 2004.

[15] Spitzer P, Zierhofer C, Hochmair E. Algorithm for multi-curve-fitting with shared parameters and a possible application in evoked compound action potential measurements. Biomedical Engineering Online 2006;5:13.

[16] Brix, G and Kiessling, F and Lucht, R and Darai, S and Wasser, K and Delorme, S and Griebel, J. Microcirculation and microvasculature in breast tumors: pharmacokinetic analysis of dynamic MR image series. Magnetic Resonance in Medicine 2004;52:420–429.

[17] Van der Meulen P, Groen J, Tinus A, Bruntink G. Fast field echo imaging: an overview and contrast calculations. Magnetic Resonance Imaging 1988;6:355–368.

[18] Sijbers J, Den Dekker A. Maximum likelihood estimation of signal amplitude and noise variance from MR data. Magnetic Resonance in Medicine 2004;51:586–594.

[19] Sourbron S, Ingrisch M, Siefert A, Reiser M, Herrmann K. Quantification of cerebral blood flow, cerebral blood volume, and blood–brain-barrier leakage with DCE-MRI. Magnetic Resonance in Medicine 2009;62:205–217.

[20] Banerji A, Caunce A, Watson Y, Rose C, Buonaccorsi G, Parker G. A flexible software phantom for generating realistic dynamic contrast-enhanced MR images of abdominal tumours. In Proc. Intl. Soc. Mag. Reson. Med, volume 16. 2008; 493.

[21] Dickie B. Joint fitting: Software to perform joint fitting to variable flip angle and DCE-MRI data, 2015. doi:10.5281/zenodo.31123.

[22] Roberts C, Little R, Watson Y, Zhao S, Buckley DL, Parker GJ. The effect of blood inflow and B1-field inhomogeneity on measurement of the arterial input function in axial 3D spoiled gradient echo dynamic contrast-enhanced MRI. Magnetic Resonance in Medicine 2011;65:108–119.

[23] Ibrahim T, Lee R, Baertlein B, Robitaille PL. B1 field homogeneity and SAR calculations for the birdcage coil. Physics in Medicine and Biology 2001;46:609.

[24] Dowell NG, Tofts PS. Fast, accurate, and precise mapping of the RF field in vivo using the 180 signal null. Magnetic Resonance in Medicine 2007;58(3):622–630.

[25] Zhang X, Petersen E, Ghariq E, Vis J, Webb A, Teeuwisse W, Hendrikse J, Osch M. In vivo blood T1 measurements at 1.5 T, 3 T, and 7 T. Magnetic Resonance in Medicine 2013;70:1082–1086.

[26] Garpebring A, Wirestam R, Östlund N, Karlsson M. Effects of inflow and radiofrequency spoiling on the arterial input function in dynamic contrast-enhanced MRI: A combined phantom and simulation study. Magnetic Resonance in Medicine 2011;65:1670–1679.

[27] Cheng HLM. T1 measurement of flowing blood and arterial input function determination for quantitative 3D T1-weighted DCE-MRI. Journal of Magnetic Resonance Imaging 2007;25:1073–1078.

[28] Chen JJ, Smith MR, Frayne R. The impact of partial-volume effects in dynamic susceptibility contrast magnetic resonance perfusion imaging. Journal of Magnetic Resonance Imaging 2005; 22:390–399.

[29] Kleppestø, Magne and Larsson, Christopher and Groote, Inge and Salo, Raimo and Vardal, Jonas and Courivaud, Frédéric and Bjørnerud, Alte. T2*-correction in dynamic contrast-enhanced MRI from double-echo acquisitions. Journal of Magnetic Resonance Imaging 2014;39(5):1314–1319.

[30] Paxton P, Curran PJ, Bollen KA, Kirby J, Chen F. Monte Carlo experiments: Design and implementation. Structural Equation Modeling 2001;8(2):287–312.

[31] Press WH, Teukolsky SA, Vetterling WT, Flannery BP. Numerical Recipes 3rd Edition: The Art of Scientific Computing. New York, NY, USA: Cambridge University Press, 2007.

[32] O'Connor JP, Rose CJ, Waterton JC, Carano RA, Parker GJ, Jackson A. Imaging Intratumor Heterogeneity: Role in Therapy Response, Resistance, and Clinical Outcome. Clinical Cancer Research 2015;21:249–257.

[33] Kershaw LE, Buckley DL. Precision in measurements of perfusion and microvascular permeability with T1-weighted dynamic contrast-enhanced MRI. Magnetic Resonance in Medicine 2006;56:986–992.

[34] Banerji A, Naish JH, Watson Y, Jayson GC, Buonaccorsi GA, Parker GJ. DCE-MRI model selection for investigating disruption of microvascular function in livers with metastatic disease. Journal of Magnetic Resonance Imaging 2012;35:196–203.

[35] Donaldson SB, West CM, Davidson SE, Carrington BM, Hutchison G, Jones AP, Sourbron SP,

Buckley DL. A comparison of tracer kinetic models for T1-weighted dynamic contrast-enhanced MRI: Application in carcinoma of the cervix. Magnetic Resonance in Medicine 2010;63:691–700.

[36] Donaldson SB, Betts G, Bonington SC, Homer JJ, Slevin NJ, Kershaw LE, Valentine H, West CM, Buckley DL. Perfusion estimated with rapid dynamic contrast-enhanced magnetic resonance imaging correlates inversely with vascular endothelial growth factor expression and pimonidazole staining in head-and-neck cancer: a pilot study. International Journal of Radiation Oncology Biology Physics 2011;81:1176–1183.

Table 1: Ground truth parameters used to generate synthetic images

Parameter (units)	Tumor Core	Tumor Rim
$T_{1,0}$ (ms)	1083	821
$M_{0,v}$ (a.u.)	10871	10500
$M_{0,d}$ (a.u.)	10871	10500
t_0 (s)	5	5
F_p (ml min^{-1} ml^{-1})	[0.10, 0.20]^a	[0.35, 0.45]^a
F_E (ml min^{-1} ml^{-1})	[0.05, 0.10]^a	[0.05, 0.10]^a
v_p (ml ml^{-1})	[0.00, 0.10]^a	[0.05, 0.15]^a
v_e (ml ml^{-1})	[0.20, 0.30]^a	[0.20, 0.30]^a

^{a}Parameters samped from uniform distributions with the given range. Ground truth values for $T_{1,0}$, $M_{0,v}$ and $M_{0,d}$ were based on measurements made in 6 patients with liver metastases (34). Ground truth 2CXM parameters for each tumor region were based on previously published analyses of clinical data (35, 36).

Table 2: Improvement in the accuracy of estimated parameters in the synthetic data

	Improvement In Accuracy (%)					
	Homogeneous B_1			Inhomogenous B_1		
	Scaled AIF (×0.5)	Unbiased AIF	Scaled AIF (×1.5)	Scaled AIF (×0.5)	Unbiased AIF	Scaled AIF (×1.5)
$T_{1,0}$	**22 (19, 25)**	**21 (20, 22)**	**22 (19, 26)**	**12 (7.0, 17)**	**10 (7.3, 13)**	**11 (6, 16)**
$M_{0,d}$	**49 (47, 51)**	**48 (47, 49)**	**48 (45, 50)**	**48 (44, 51)**	**46 (44, 48)**	**46 (43, 50)**
t_0	**14 (9.3, 18)**	**8.7 (6.4, 11)**	**8.3 (3.2, 13)**	**7.7 (0.36, 15)**	2.0 (-2.8, 6.5)	1.5 (-6.3, 8.8)
F_p	**15 (11, 18)**	**11 (9.1, 12)**	-5.4 (-9.6, -1.4)†	**21 (16, 26)**	**17 (15, 20)**	2.6 (-3.1, 8.0)
F_E	**10 (6.5, 13)**	**1.8 (0.087, 3.5)**	-8.3 (-13, -4.3)†	**12 (6.8, 17)**	**3.7 (0.38, 7.0)**	-6.2 (-12, -0.44)†
v_p	**10 (6.7, 13)**	**13 (11, 14)**	-0.63 (-4.5, 3.1)	-7.1 (-13, -1.5)†	-3.8 (-7.3, -0.34)†	-20 (-27, -14)†
v_e	**3.7 (0.16, 7.2)**	**7.7 (6.1, 9.2)**	-5.9 (-9.8, -2.1)†	**5.5 (0.34, 10)**	**9.3 (6.3, 12)**	-4.0 (-9.6, 1.4)

Percentage improvement in accuracy of parameter estimates for joint fitting versus sequential fitting. Improvements in accuracy were computed for each experimental condition from the linear model coefficients. Values in parentheses are Bonferroni corrected 95% confidence intervals. Bold denotes statistically signficant improvements in accuracy compared to sequential fitting. Negative values represent degradation in accuracy. Dagger denotes statistically significant degradation in accuracy.

Table 3: Improvement in the precision of estimated parameters in the synthetic data

Improvement In Precision (%)

	Homogeneous B_1			Inhomogenous B_1		
	Scaled AIF (×0.5)	Unbiased AIF	Scaled AIF (×1.5)	Scaled AIF (×0.5)	Unbiased AIF	Scaled AIF (×1.5)
$T_{1,0}$	**20 (17, 23)**	**20 (19, 22)**	**20 (17, 23)**	**21 (17, 26)**	**21 (18, 24)**	**21 (17, 26)**
$M_{0,d}$	**48 (46, 50)**	**48 (46, 48)**	**47 (45, 50)**	**50 (47, 53)**	**50 (48, 53)**	**50 (46, 53)**
t_0	**6.8 (3.9, 9.5)**	**8.1 (6.9, 9.4)**	**9.5 (6.7, 12)**	**5.4 (1.3, 9.4)**	**6.8 (4.3, 9.3)**	**8.2 (4.1, 12)**
F_p	**15 (11, 18)**	**12 (11, 14)**	**14 (11, 17)**	**17 (13, 22)**	**15 (12, 18)**	**17 (12, 21)**
F_E	**5.1 (1.5, 8.7)**	-0.91 (-2.6, 0.79)	1.7 (-2.1, 5.4)	**6.4 (1.1, 11)**	0.45 (-3.0, 3.8)	3.1 (-2.4, 8.2)
v_p	**18 (15, 21)**	**18 (16, 19)**	**17 (14, 20)**	**19 (14, 23)**	**19 (16, 22)**	**18 (13, 22)**
v_e	**10 (6.6, 13)**	**4.7 (3.0, 6.3)**	3.2 (-0.51, 6.8)	**5.7 (0.40, 11)**	0.024 (-3.4, 3.3)	-1.5 (-7.2, 3.9)

Percentage improvement in precision of parameter estimates for joint fitting versus sequential fitting. Improvements in precision were computed for each experimental condition from the linear model coefficients. Values in parentheses are Bonferroni corrected 95% confidence intervals. Bold denotes statistically signficant improvements in precision compared to sequential fitting. Negative values represent degradation in precision. Dagger denotes statistically significant degradation in precision.

Table 4: Improvement in the precision of estimated parameters in the clinical data

	Improvement In Precision (%)		
	Scaled AIF (×0.5)	Measured AIF	Scaled AIF (×1.5)
$T_{1,0}$	-8.5 (-15, -2.4)†	-4.9 (-7.5, -2.4)†	-4.0 (-10, 1.9)
$M_{0,d}$	**37 (34, 41)**	**38 (36, 39)**	**38 (35, 41)**
t_0	**9.4 (5.7, 13)**	**6.5 (4.9, 8.0)**	**4.6 (0.67, 8.3)**
F_{p}	**22 (18, 26)**	**22 (20, 23)**	**22 (18, 26)**
F_{E}	**11 (6.9, 16)**	**13 (11, 14)**	**11 (6.9, 16)**
v_{p}	**16 (11, 20)**	**17 (16, 19)**	**16 (12, 20)**
v_{e}	**14 (8.8, 18)**	**15 (13, 16)**	**13 (8.4, 18)**

Percentage improvement in precision of parameter estimates for joint fitting versus sequential fitting. Improvements in precision were computed for each experimental condition from the linear model coefficients. Values in parentheses are Bonferroni corrected 95% confidence intervals. Bold denotes statistically significant improvements in precision compared to sequential fitting. Negative values represent degradation in precision. Dagger denotes statistically significant degradation in precision.

Figure Captions

1. Synthetic data used for evaluation of sequential and joint fitting methods. Figure part (a) shows the central slice of the full phantom. Figure parts (b) and (c) show zoomed images of the synthetic dynamic data without and with simulated B_1 field inhomogeneities respectively.

2. Arterial input functions (a and b), example Monte Carlo and residual bootstrap fits (c and d) and densities for each estimated parameter (e and f). The left column (a, c, and e) show data from the synthetic tumor. The right column (b, d, and f) shows data from a randomly selected tumor in the main clinical experiment. Figure parts (c) and (d) show all 100 Monte Carlo or residual bootstrap fits for representative voxels taken from synthetic and bladder tumors respectively. Parts (e) and (f) show corresponding densities for parameter estimates obtained from fits shown in the central row of (c) and (d). In (e), black vertical lines represent ground truth.

3. Parametric maps for $T_{1,0}$, $M_{0,d}$, F_p, v_p, F_E and v_e obtained using sequential and joint estimation for an example slice of the synthetic tumor. In the homogeneous B_1 field condition, jointly estimated $M_{0,d}$ maps appear smoother than their sequential estimation counterparts. Differences in the appearance for the other estimated parameter maps are difficult to discern visually. Under the inhomogeneous B_1 field condition, clear visual differences arise in the maps of F_p and v_p, with joint estimation maps showing less speckle. The reduction of speckle observed in the F_p map enables rim-core subregions present in the ground truth map to be more clearly identified.

4. Parametric maps for $T_{1,0}$, $M_{0,d}$, F_p, v_p, F_E and v_e obtained using sequential and joint estimation for two example tumors. On average, estimates of $T_{1,0}$ and $M_{0,d}$ appear higher when using joint estimation (higher $T_{1,0}$ and $M_{0,d}$ at the top of tumor 1 and center of tumor 2). In tumor 1, joint estimates of F_p appear less sensitive to errors in the AIF, and v_p maps are less speckled than their sequential estimation counterparts. In tumor 2, spurious values of v_p, F_E, and v_e observed in the sequential maps occur less frequently with joint estimation.

Supporting Figure Captions

1. Comparison of $T_{1,0}$ estimates in prostate tissue obtained using sequential and joint estimation in VFA and DCE-MRI data against estimates obtained using an independent inversion-recovery turbo field echo (IR-TFE) sequence, which was taken as gold-standard. Both sequential and joint estimation overestimated $T_{1,0}$ relative to IR-TFE measurements, however joint estimation led to a

reduction in the mean relative error of 7%.

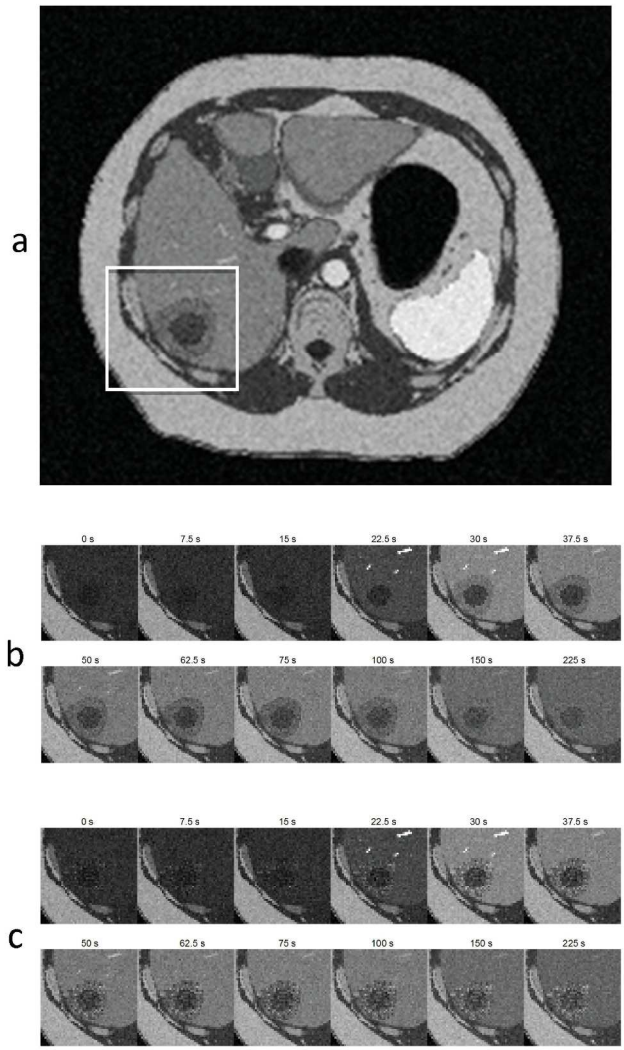


Figure 1. Synthetic data used for evaluation of sequential and joint fitting methods. Figure part (a) shows the central slice of the full phantom. Figure parts (b) and (c) show zoomed images of the synthetic dynamic data without and with simulated B_1 field inhomogeneities respectively.
200x307mm (300 x 300 DPI)

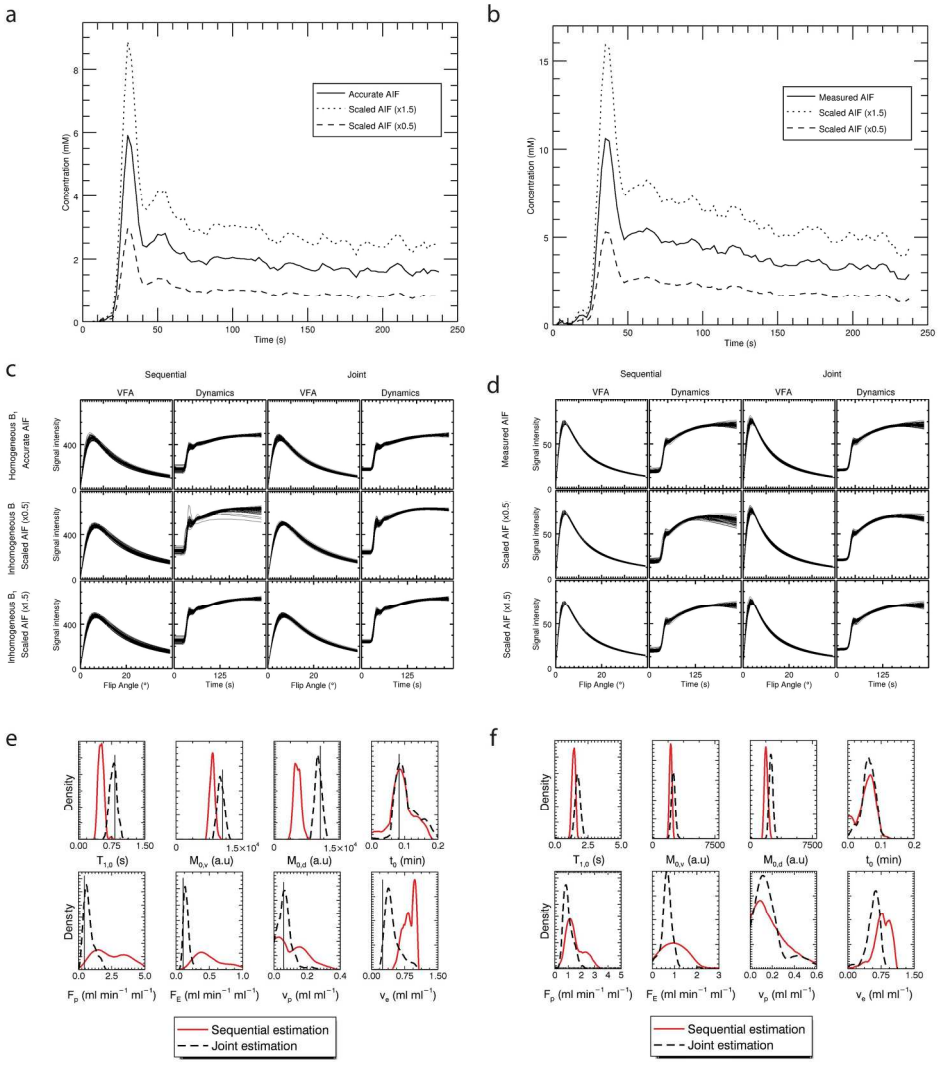


Figure 2. Arterial input functions (a and b), example Monte Carlo and residual bootstrap fits (c and d) and densities for each estimated parameter (e and f). The left column (a, c, and e) show data from the synthetic tumor. The right column (b, d, and f) shows data from a randomly selected tumor in the main clinical experiment. Figure parts (c) and (d) show all 100 Monte Carlo or residual bootstrap fits for representative voxels taken from synthetic and bladder tumors respectively. Parts (e) and (f) show corresponding densities for parameter estimates obtained from fits shown in the central row of (c) and (d). In (e), black vertical lines represent ground truth.
196x221mm (300 x 300 DPI)

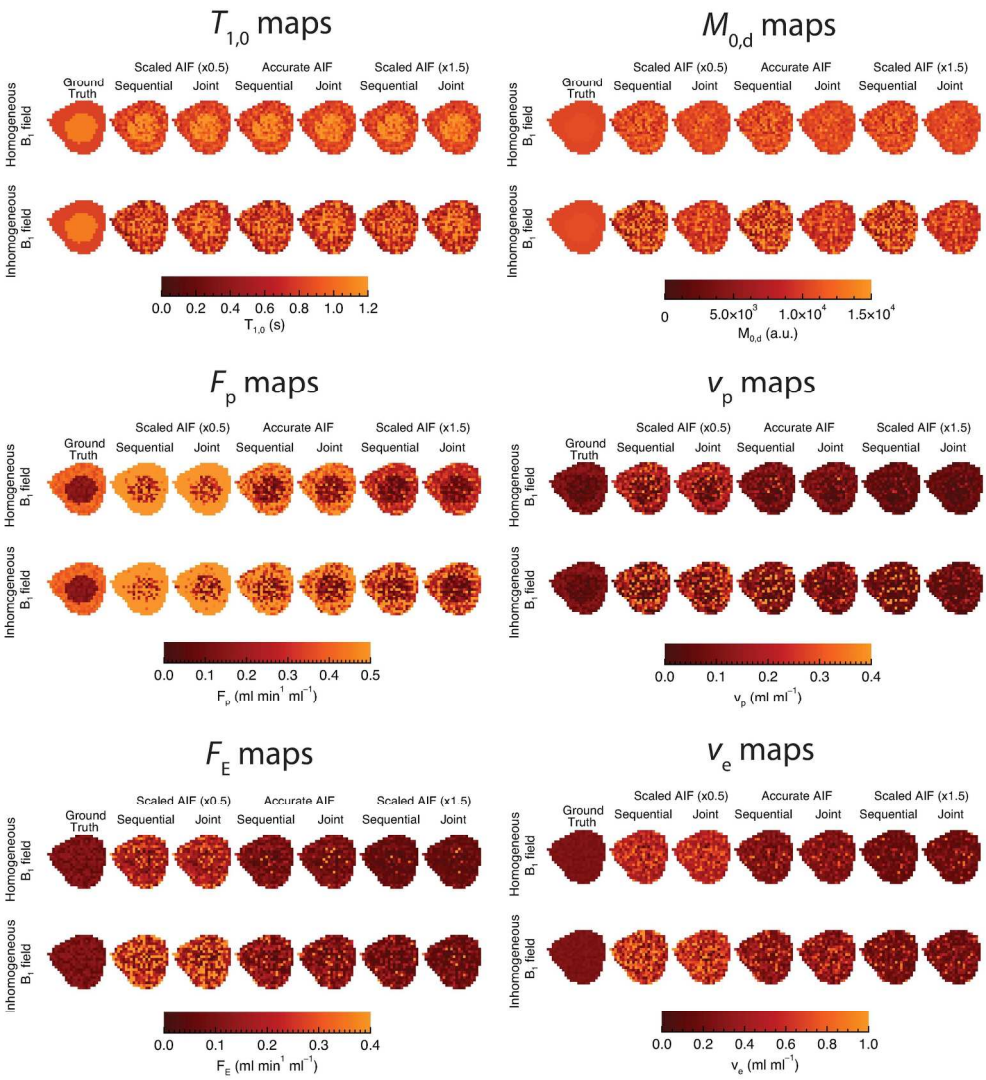


Figure 3. Parametric maps for $T_{1,0}$, $M_{0,d}$, F_p, v_p, F_E and v_e obtained using sequential and joint estimation for an example slice of the synthetic tumor. In the homogeneous B_1 field condition, jointly estimated $M_{0,d}$ maps appear smoother than their sequential estimation counterparts. Differences in the appearance for the other estimated parameter maps are difficult to discern visually. Under the inhomogeneous B_1 field condition, clear visual differences arise in the maps of F_p and v_p, with joint estimation maps showing less speckle. The reduction of speckle observed in the F_p map enables rim-core subregions present in the ground truth map to be more clearly identified.
192x218mm (300 x 300 DPI)

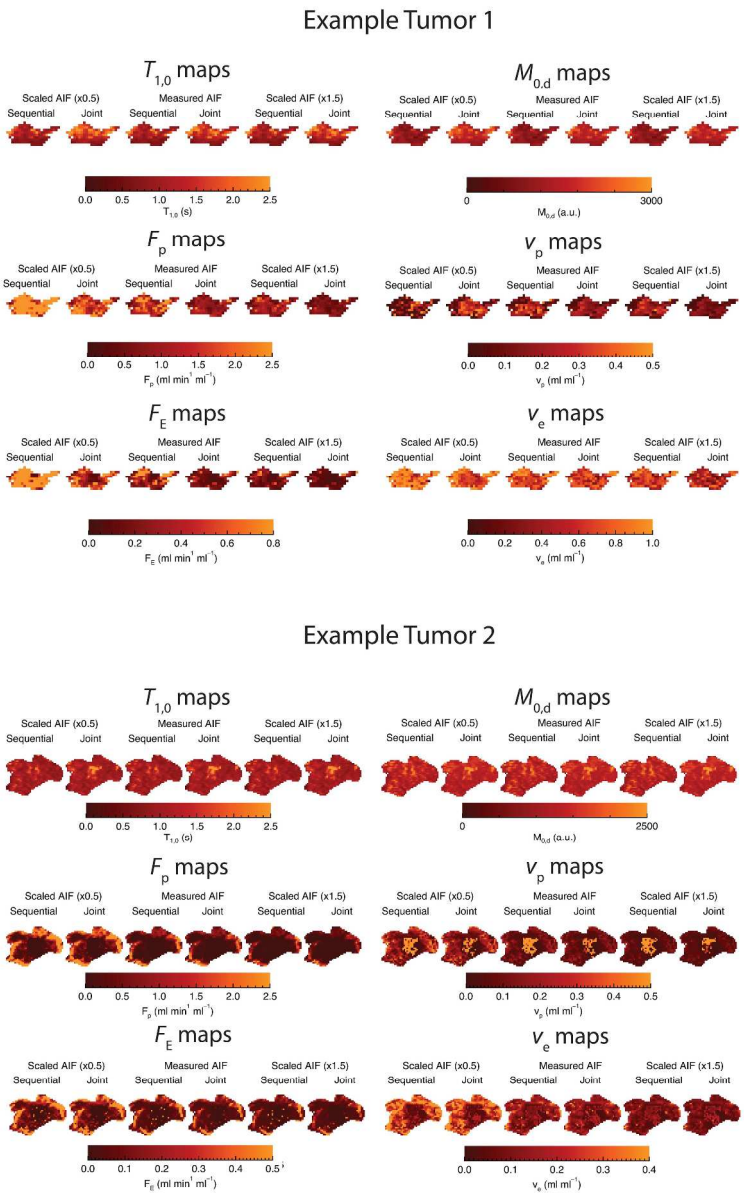


Figure 4. Parametric maps for $T_{1,0}$, $M_{0,d}$, F_p, v_p, F_E and v_e obtained using sequential and joint estimation for two example tumors. On average, estimates of $T_{1,0}$ and $M_{0,d}$ appear higher when using joint estimation (higher $T_{1,0}$ and $M_{0,d}$ at the top of tumor 1 and center of tumor 2). In tumor 1, joint estimates of F_p appear less sensitive to errors in the AIF, and v_p maps are less speckled than their sequential estimation counterparts. In tumor 2, spurious values of v_p, F_E, and v_e observed in the sequential maps occur less frequently with joint estimation.
284x396mm (300 x 300 DPI)

Supporting Materials

Comparison of sequential and joint VFA $T_{1,0}$ estimates with reference measurements

Introduction

The variable flip angle (VFA) method is commonly used in dynamic contrast-enhanced (DCE) MRI examinations because it facilitates rapid estimation of pre-contrast T_1 ($T_{1,0}$). $T_{1,0}$ is required to estimate contrast agent concentrations and hence facilitate the fitting of tracer kinetic models to DCE-MRI data. However, it is known that the VFA method is less accurate and precise than other $T_{1,0}$ mapping methods. Here we test the hypothesis that our proposed joint estimation method improves $T_{1,0}$ accuracy compared to the conventional sequential estimation approach, using independent measurements of $T_{1,0}$ made using an inversion-recovery turbo-field echo (IR-TFE) sequence in a clinical prostate cancer study. IR-TFE measurements are more accurate than VFA measurements of $T_{1,0}$, and were therefore taken as gold-standard for this experiment.

Method

Analysis was performed using data from an ongoing clinical prostate cancer study being performed at our center. At the time of analysis, analyzable data from 3 patients had been acquired. All patients gave written informed consent and approval was obtained from the local research ethics committee. All imaging was performed on a 1.5 T Philips Achieva MR scanner. High spatial resolution T_2-w imaging was performed to allow accurate delineation of the prostate in all patients. This was followed by VFA, IR-TFE and DCE-MRI examinations, acquired at a lower spatial resolution ($2.3 \times 2.3 \times 5.0$ mm^3), for assessment of sequential and joint $T_{1,0}$ accuracy. Field of view was matched between all examinations, and matrix sizes were matched between VFA, IR-TFE and DCE-MRI acquisitions. Acquisition parameters for the IR-TFE sequence were: TR/TE of 2.38/0.77 ms; shot interval of 4000 ms; matrix size of 176 x 176 x 20; flip angle of 12°; and inversion times of 64 ms, 250 ms, 1000 ms, 2500 ms, and 3900 ms. Acquisition parameters for the VFA and dynamic SPGR sequences were: TR/TE of 2.47/0.86 ms; variable flip angles of 2°, 10° and 20°; NSA in the VFA data of 5, dynamic flip angle of 30°; NSA in the dynamic data of 1, dynamic temporal resolution of 1.6 s; and dynamic acquisition time of 6.8 minutes. SENSE factors of 2.5 were used for VFA, IR-TFE and dynamic imaging. Sequential and joint estimates of $T_{1,0}$ were generated in an identical manner to that described in the main paper.

The null hypothesis of no difference in mean relative error between sequential and joint estimates of $T_{1,0}$ (relative to IR-TFE measurements) was tested using a paired two-sided t-test with significance criterion $P < 0.05$. A paired test was used to account for correlation between sequential and joint

R3.2 removal of statement of independence

estimates made at the same voxel.

Statistical analysis was performed at the voxel level ($n = 1532$) giving a power of 99% to detect a difference in mean relative error of 10% assuming a standard deviation of 50%. Analyses were performed in R (Version 3.1, R Foundation for Statistical Computing, Vienna, Austria). Bland Altman plots were generated to show the relative error in $T_{1,0}$ (sequential/joint $T_{1,0}$ minus IR-TFE $T_{1,0}$) across the range of measured gold-standard $T_{1,0}$ values. R3.1 clarification of direction bias

Results

Acquisition time for VFA data was 24 s. Acquisition time for IR-TFE data was 4 minutes. Supporting Figure 1 shows Bland Altman plots of the relative error in $T_{1,0}$ for sequential and joint estimates. Both sequential and joint estimation overestimated $T_{1,0}$ relative to IR-TFE measurements. While joint estimation caused a small number of estimates at low $T_{1,0}$ values (between 1.0 s and 1.5 s) to have larger relative error than sequential estimates (shown as a negative shift from zero error), a far greater number of estimates in the same $T_{1,0}$ range were shifted from positive relative error towards zero relative error, leading to an overall reduction in the mean relative error. Mean relative error for sequential and joint estimates of $T_{1,0}$ were 35% and 28% respectively, corresponding to a reduction of 7.0% (95% CI 5.4-8.5%, P-value < 0.0001). Using the same terminology as the main paper, joint estimation therefore led to an improvement in the mean relative error of 20% (95% CI 15-24%, P-value < 0.0001), which agrees well with results from the synthetic experiment (improvements of between 10-22%). R3.1 clarification of direction bias

Equality of $M_{0,v}$ and $M_{0,d}$

Introduction

Joint fitting of the equilibrium longitudinal magnetization relies on the assumption that $M_{0,v} = M_{0,d}$. This condition is satisfied only if the MR scanner uses the same receive gain settings for variable flip angle and dynamic imaging. Although steps were taken to avoid recalibration, DICOM attributes for scanner receive gains were checked for equality (e.g. aFFT.SCALE.n.flFactor on Siemens, where n denotes the receive channel), and phantom experiments to estimate $M_{0,v}$ and $M_{0,d}$ within variable flip angle and dynamic images were performed. The phantom experiment for the bladder protocol is described in detail below.

Method

A commercial uniformity phantom (Eurospin II $T01$ Flat Field Phantom, Diagnostic Sonar Ltd., Livingston, Scotland) was filled with copper sulphate solution and scanned using the bladder DCE-MRI protocol described in the main paper. A circular region of interest (ROI) with diameter 0.9 times that

of the phantom was placed in the central slice of each variable flip angle image. The equilibrium longitudinal magnetization was estimated for variable flip angle images i and ROI voxels j using the following steady-state SPGR equation:

$$M_{0,v}(i,j) = \frac{S(i,j)(1-\cos(\theta_{v_i})e^{-\frac{TR}{T_{1,0}}})}{\sin\theta_{v_i}(1-e^{-\frac{TR}{T_{1,0}}})} \quad (1)$$

where $T_{1,0}$ was assumed to be 850 ms based on measurements described in (1). All other symbols were defined in the main paper. It was assumed there was no calibration between acquisition of successive variable flip angle images (i.e. $M_{0,v}(1,j) = M_{0,v}(2,j) \ldots M_{0,v}(n,j)$), and therefore the mean and standard deviation of $M_{0,v}$ was calculated across all i and j.

For the dynamic images a similar procedure was performed. The equilibrium longitudinal magnetization was estimated for all ROI voxels j at the 10th dynamic time point. A late time point was chosen to ensure steady state conditions:

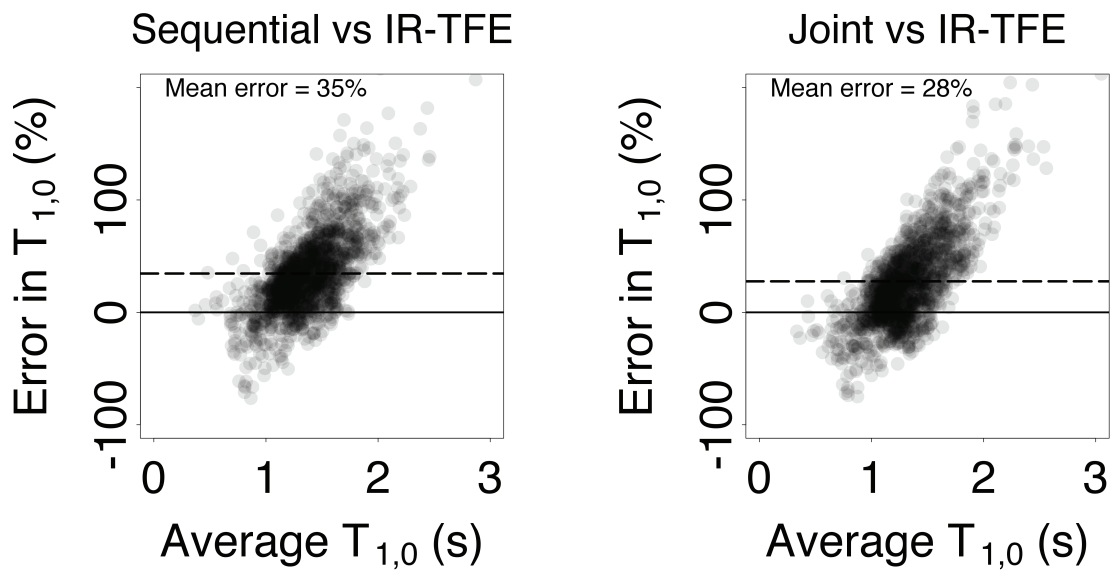
$$M_{0,d}(10,j) = \frac{S(10,j)(1-\cos(\theta_d)e^{-\frac{TR}{T_{1,0}}})}{\sin\theta_d(1-e^{-\frac{TR}{T_{1,0}}})} \quad (2)$$

the mean and standard deviation of $M_{0,d}$ was calculated across all j.

Results

The mean and standard deviation for $M_{0,v}$ and $M_{0,d}$ were found to be $(12.6 \pm 2.8) \times 10^3$ a.u. and $(12.8 \pm 2.7) \times 10^3$ a.u. respectively. A two-tailed Students t-test was performed in R (version 3.1) to test the null hypothesis of no difference between $M_{0,v}$ and $M_{0,d}$. A P-value of 0.37 was observed, confirming our belief that the scanner did not recalibrate between variable flip angle and dynamic imaging.

Supporting Figures



Comparison of $T_{1,0}$ estimates in prostate tissue obtained using sequential and joint estimation in VFA and DCE-MRI data against estimates obtained using an independent inversion-recovery turbo field echo (IR-TFE) sequence, which was taken as gold-standard. Both sequential and joint estimation overestimated $T_{1,0}$ relative to IR-TFE measurements, however joint estimation led to a reduction in the mean relative error of 7%.

Bibliography

[1] Lerski R, McRobbie D. Eurospin II Magnetic Resonance Quality Assessment Test Objects. Diagnostic Sonar Ltd, 1992.